

Voice and graphical -based interfaces for interaction with a robot dedicated to elderly and people with cognitive disorders

C. Granata, M. Chetouani, A. Tapus, P. Bidaud, V. Dupourqué

Abstract—Human-robot interaction (HRI) takes place especially through interfaces. The design of such interfaces is a very delicate and crucial phase because it influences the robot accessibility and usability by the user. In this paper, we describe and analyze the results of 2 tests conducted so as to understand some of the optimal features that should characterize the robot voice and graphical-based user interfaces. Our test platform is an assistive robot developed for the elderly with mild cognitive impairments. Therefore, the user interfaces must be clear and simple. The ambiguities must be eliminated so as to facilitate the use of the robot and hence not to discourage the elderly population to use new technologies.

I. INTRODUCTION

The world's population is growing older. Most of the elderly population suffers from the effects of social isolation and age related cognitive decline (deterioration in memory, attention, concentration, etc) [1], [2]. The new trend in assistive technologies is promoting aging of elderly persons at home. It is still necessary to define the skills and the characteristics which a social robot should have. The authors in [3] describe a social robot capable of interacting and communicating with humans by following the behavioral norms expected by the people. Hence, having a social robot capable of moving and acting independently and in the same time capable of assisting people in their own home environments is still a challenge.

An important question which it is necessary to be addressed is to know if an elderly and/or a non-expert user would be able to interact naturally and intuitively with social robots. Voice interactions are the most common way people express their needs. Voice based interfaces are more powerful than graphical based ones. However, the population to whom the robot is dedicated has cognitive, memory, and/or understanding problems. Gödde et al. [4] have shown that profiling user input is particularly important for the elderly population. The authors have compared the interactions of older and younger users with a speech-based smart home

Consuelo Granata, PhD student of ISIR (Université Pierre et Marie Curie-Paris6-UPMC) and Robosoft (Bidart) consuelo.granata@robosoft.com

Mohamed Chetouani Institut des Systèmes Intelligents et Robotique (ISIR), Université Pierre et Marie Curie-Paris6 (UPMC), 4 Place Jussieu, 75252 Paris Cedex, France mohamed.chetouani@upmc.fr

Adriana Tapus, Computer Science Department Ecole Nationale Supérieure de Techniques Avancées (ENSTA-ParisTech) 32 Blvd Victor, 75015, Paris, France adriana.tapus@ensta.fr

Philippe Bidaud Institut des Systèmes Intelligents et Robotique (ISIR), Université Pierre et Marie Curie-Paris6 (UPMC), 4 Place Jussieu, 75252 Paris Cedex, France philippe.bidaud@upmc.fr

Vincent Dupourqué, Robosoft, Technopole d'Izarbel, 64210 Bidart, France vincent.doupourque@robosoft.com



Fig. 1. The mobile robot Kompai (by Robosoft -France) used for interaction.

system. The results illustrated that older users were less likely to speak to the system in a way that was easy for the system to understand, and achieved lower task success. Age-related changes affect many interrelated aspects of cognition, such as information processing speed, mental flexibility, fluid intelligence, and memory [5]. The memory deficits in the elderly impair language comprehension and production. The interfaces used should be able to handle potential problems of human speech such as sentence fragments, false starts and interruption.

To allow the elderly with cognitive impairments to interact comfortably with the machine, we propose a multimodal approach: a combination of voice and graphical support. A classical voice-based interface is composed of an Automatic Speech Recognizer (ASR) that interprets the human voice, a Dialogue Manager that selects the appropriate response, and a Text-To-Speech (TTS) module that synthesizes the selected response. As showed in Fig 2, we propose a multimodal approach adding the fusion process to combine voice and graphical inputs and the fission process to combine voice and graphical output.

Our aim is to create a graphical support that accompanies the voice interaction. To facilitate the use of the robot, the system of fission must be designed according to some well founded criteria concerning the language, the vocabulary (words have to be chosen with respect to their frequency and familiarity), the syntax (the rules by which words are organized to form a correct sentence should be simple), and semantics (the meaning of a word or of a sentence should

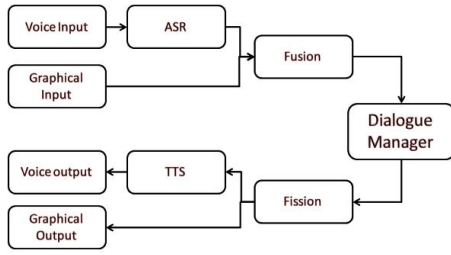


Fig. 2. Multimodal approach using the fusion and fission process to combine voice and graphical input/output.

avoid any ambiguities). However, if for the linguistic part there are some solid rules that can be employed so as to define a model, for the graphical-based interface, substantiated criteria are still not very well known. In this paper, we posit that the concomitant use of voice and graphical support can increase the usability of the robot by the elderly and people with mild cognitive impairments.

II. RELATED WORKS

Different kinds of interfaces have been tested over the past few years to allow a non-expert human to interact with a robot. Some tried to address this challenge by transposing human/human modes of interaction based on voice recognition, gesture recognition and gaze tracking [6], [7], [8]. Other interfaces have been developed using some handheld devices such as the Wiimote [9], the Personal Digital Assistant (PDA) [10] and the I-Phone [11]. In order to enhance the flexibility and the naturalness of the user interfaces, others researchers have developed multimodal systems. Reichman [12], [13] proposed the conversation metaphor, a hybrid interaction style, which integrates graphical and user's natural language. In [14], the authors explain a multimodal interface combining speech recognition, hand gesture recognition, and inputs from a PDA to drive mobile robots. Concerning language in the elderly, relatively little has been published on fission process in HRI interaction. Wolters et al. [15] assessed two approaches to accommodate users with memory troubles: on one hand they wanted to reduce the number of options proposed by the machine and on the other hand the system was supposed to provide confirmations (if the system confirms each aspect of the interaction, users will find it easier to remember it). At the cognitive level, we know that there is a relation between an image and the linguistic representation of an object (or of a scene) represented in an image. In recent years, many researches concerning the graphical representation of objects have been conducted. Snodgrass and Vanderwart [16] were the pioneers towards a phase of standardization. Indeed, these authors have established an extended set of pictures (black-and-white line drawings) that have been standardized on four variables (name agreement, image agreement, familiarity, and visual complexity). In our study, we made reference to the studies of Alario and Ferrand [17] (see Fig 3), Bonin and Peereman [18], and to some well-defined databases such as Picture Communication Symbols

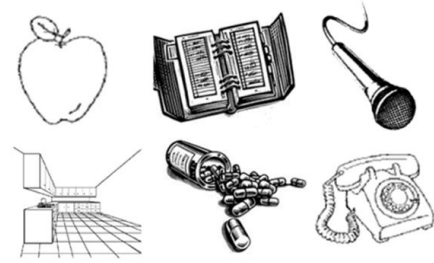


Fig. 3. Some images from the Alario&Ferrand and Lexique databases [22], [23].



Fig. 4. 1,6 images from the BoardMaker database; 2, 3, 4, 5 images from the Fotolia database.

(PCS) [19], BoardMaker [20], and Fotolia [21] (see Fig 4) in order to select the pictures used in our graphical interface. We think that the careful selection of sentences/words uttered by the robot and of the images displayed on the Tablet PC can facilitate the understanding of the robot's behavior by the user.

III. ROBOT TEST BED

The tests proposed in this paper are based on a static interaction between the user and a robot. They can be made also with a simply computer rather than a mobile robotics platform. However, we have chosen the same to using our robot in order to not condition the participant's response. In fact the approach of people with a simple computer is totally different from the approach with a mobile robot, especially for elderly people.

A. Hardware

The robot used in this work is a Kompaï robot (Fig 1), a newly developed companion robot. The robot embeds all necessary sensors for autonomous navigation:

- A laser sensor for localization and obstacle detection
- Ultrasound and Infrared sensors for obstacle detection
- 2 Cameras

The low-level control is managed by the embedded Microcontroller (Emtrion SH7780 SBC). For high-level control and user interfaces, the Kompaï uses a Tablet PC running Windows Vista.

B. Software

The robot's functionalities currently accessible by the voice-based interface are:

- Hear management (start/stop robot listening)

Function	Number of keywords	Number of syntax models	Number of showed images + buttons
Date and hour	3	10	0
Shopping	10	14	54
Appointment	20	18	6
Wake up	5	2	6
Medical	9	2	7
Make robot moving	5	3	30
E-mails	8	6	8
Telephone	4	6	8
Listen start/stop	3	3	7

Fig. 5. Number of keywords and syntactic models defined in our vocal-based interface for each robot function. In the third column, number of images and buttons proposed for graphical-based interface.

- Shopping list, appointment, and drugs management
- General information (time and date)
- Wake up
- Robot movement
- Send e-mails
- Medical diagnostic

The management of shopping list, agenda, table of drugs and e-mails sending is made using Google API developed for .NET application. The robot movements exploit a SLAM (Simultaneous Localization and Mapping) software. The SLAM software used is based on the Karto®robotic navigation software under license from SRI International®. Our dialogue system consists of 3 different components: 1) ASR which converts audio signals of human speech into text strings and matches the results of the analysis with the known vocabulary and syntax; 2) Dialogue Manager which maps the meanings and intentions of the recognized utterances, and selects the appropriate response; 3) TTS which converts the system utterances into speech output. The Dialogue application runs under Windows Vista OS and is programmed in MRDS (Microsoft Robotics Dev Studio) platform.

IV. METHODOLOGY

A. Voice-based interface

Concerning the voice-based interface an important research question must be answered: "Which words and which syntax the robot has to use to be easily understood by the elderly?" We posit that in order to answer to this question we have to study the speaking habits of the elderly and of people suffering of mild cognitive troubles. If we understand some of these speaking patterns, we will be able to appropriately model the robot way to speak to the users. In order to build our voice-based interface, we first analyze the type of vocabulary to be employed by the users, the frequency of the words, and the syntax and semantics of sentences used. The choice of the words and of the sentence structure will be made considering the results of this work. Our current system contains a vocabulary of about 170 french words. We have selected these words with respect to the robot capabilities and functionalities exposed in Section III. Additionally to the robot functionalities developed for our application, we have also built different syntactic models. Such an example is described below: in order to ask to the

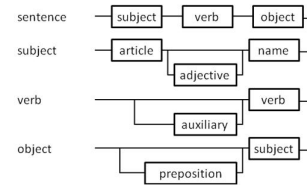


Fig. 6. An example of simple syntactic model in French.

robot the date, the user can say different sentences ("Quel jour sommes-nous aujourd'hui?"¹, "S'il vous plait, quelle est la date d'aujourd'hui?"²). In the table on Fig 5 the number of words and syntactic models defined for the different robot functionalities are reported. In this table not all the words of the defined vocabulary are enumerated (only the keywords). Taking as an example the Shopping function, some defined keywords are "Ajouter", "Ecrire", "Supprimer", and "Effacer"³. The words as "Pomme", "Fromage", and "Tomate"⁴ are not considered as keywords. The idea is that the user can also add to the system the names of products if they are not defined in the grammar, using the graphical-based interface. In order to help users to understand the sentences pronounced by the robot, we worked closely with our colleagues from Pitie-Salpetriere Hospital, specialists in psychology and cognitive neurosciences. We have learnt that the robot sentences do not have to be too long or too complex. The syntactic structure must be simple in order to avoid to the elderly a cognitive overload. One way to categorize sentences is by the clauses they contain. A simple sentence contains a single, independent clause (ex: Today you have an appointment with the doctor"); a compound contains two independent clauses that are joined by a coordinating conjunction (ex: "Today you have to meet the doctor and to call your son"); a complex sentence contains an independent clause plus one dependent clause (ex: "Today you have to meet the doctor who called you yesterday"). The language of the robot has been chosen on the basis of simple syntactic models. Fig 6 is depicting a detailed example of a simple french sentence.

B. Graphical-based interface

When we started focusing on the graphical-based interface an important research question arose: "Which type of image has to be used to make the graphical-based robot interface easily understandable and, in the same time, motivating for the elderly?". Moreover, we wanted to understand the style of image to use in the graphical-based interface for robots dedicated to the elderly: simple or rather complex, user-friendly, drawings or photos, image with or without text, colored or black-and-white images. We couldn't find the answers to these questions in the literature, considering that our image selection is specifically oriented to service robotic applications for elderly. We have consulted two existing

¹ "What day is it today? "

² "Please, what's today's date? "

³Add, Write, Delete, Erase

⁴Apple, Cheese, Tomato

databases ([22], [23]) in order to select the desired images. Each image of these databases is considered by the literature as the most representative for a specific word for the majority of people. After defining the robot functionalities and services accessible from the graphical-based interface (the same functions also accessible from the voice-based interface), we have collected the related pictures. Moreover, we have also searched in other image databases (i.e., BoardMaker and Fotolia) some very similar and some very different pictures, including both drawings and photos, both colored and black-and-white images, both simple and complex images (see an example in Fig 7). We seek the best match image-word; this means that we tried to understand which type of image the patient associated as easy as possible to a given word.

V. EXPERIMENTAL DESIGN

A. Subject Pool

In order to test and validate our methodology, 11 patients were recruited from our project partner the Memory Clinic of Broca geriatric hospital in Paris (France). 6 patients participated to the tests concerning the voice-based interface: 5 females and 1 male; 3 MCI⁵, and 3 normal people all aged between 66 and 79 years. 5 patients participated to the tests concerning the graphical-based interface: 3 females and 2 males; 2 MCI and 3 normal people all aged between 67 and 74 years. The patients have signed the Informed Consent and the entire intervention has been recorded on video tape.

B. Voice-based interface - Experimental Design

First of all, we have explained to the patients the role of the robot in their everyday life. Furthermore, each patient had to achieve the following tasks / functions with the robot:

- Ask for date and time
- Ask the content of the shopping list
- Add something on the shopping list
- Remove something from the shopping list
- Request to wake up
- Request the appointments of the day
- Make an appointment with somebody
- Ask for the next appointment with somebody

With this test session, we wanted to know the syntax models and the words used frequently by the targeted patients to interact with the robot. This test was done using the "Wizard of Oz" framework.

C. Graphical-based interface - Experimental Design

For each set of proposed images, the patients had to express their preference from different points of view:

- 1) Which is the most consistent image to represent the object?
- 2) Which is the most convivial (user-friendly) image?
- 3) Which is the image that they would like on the robot graphical-based interface?

⁵MCI=Mild Cognitive Impairment

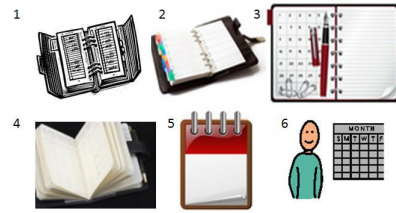


Fig. 7. For the function Agenda 1. Image of the Alario-Ferrand database; 2, 3, 4, 5. Chosen images from Fotolia; 6. Image selected from BoardMaker.

We have also proposed to the patients to write 3 words ("Bonjour"⁶, "Soleil"⁷, and "Village"⁸) with a keyboard displayed on the touch screen. We have repeated this test twice. The first time, we have proposed an azerty keyboard layout and, the second time we have proposed an alphabetical keyboard layout (i.e., the letters are placed in alphabetical order). The main goal was the comparison of the usability of the two different keywords.

VI. EXPERIMENTAL RESULTS

A. Voice-based interface

In order to understand the syntactic models and the vocabulary used by patients, we have annotated and analyzed all the sentences uttered. After analysis (annotation from videos recorded during the tests), we realized that the syntactic models and vocabulary used by elderly are totally different from those used by non-elderly adults (e.g. the use of courtesy words and the modes of expression differ considerably). For example, the most part of the young/adult French population, ask the scheduling of the day as follows: "Est-ce que j'ai des rendez-vous aujourd'hui?"⁹. The elderly say most formal sentences as "Pourriez-vous m'indiquer si j'ai des rendez-vous aujourd'hui?"¹⁰. Between all the sentences uttered by all patients, only 44.5% coincides exactly with the sentences and with the words known by the robot, and hence used by the robot, but only 13.4% of sentences differ from the syntactic structure known by the Dialogue Manager. So, as it was predictable, the patients use very linear and simple syntax, close to the model showed in Fig 6. Concerning the vocabulary, 42.1% of uttered sentences contain some words unknown to the robot. This means that the robot language and the elderly language don't coincide. We can conclude that in order to facilitate the use of our system to the elderly, we have to enrich the vocabulary with words employed by the target population. Moreover, in order to ameliorate further the voice-based interface we also need to add others simple syntactic models to our existing model.

B. Graphical-based interface

For our tests, we have presented 24 different sets of images to the users. Each set represents a service or an object

⁶"Good morning"

⁷"Sun"

⁸"Village"

⁹" Do I have any appointment today? "

¹⁰" Can you tell me if I have any appointment today? "

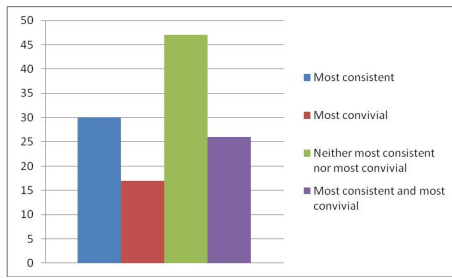


Fig. 8. The set of images preferred for the robot interface. The majority of these images are considered neither as the most consistent nor as the most convivial.



Fig. 9. For the function Shopping 1. Big icon without text; 2. Smaller icon including the object name.

concerning the robot. For each of these sets, each patient has chosen: 1. the most consistent image; 2. the most convivial (user-friendly) image; 3. the most adequate image for the robot graphical-based interface. Of course, for each patient the 3 choices can agree or be different. We have 120 images chosen as the most adapted to be on the robot interface (24 sets of images for 5 patients). From these 120 images, the following results have been obtained: 30 images have been chosen by the majority of people as the most consistent, 17 images have been selected as the most convivial, and 26 images have been considered at the same time the most consistent and the most convivial. The other 47 images chosen as the most adapted for the graphical-based interface have been considered neither the most consistent, nor the most convivial. This means that neither the consistency nor the conviviality is the adequate criteria to select good images for the interface (Fig 8).

Analyzing the images preferred for the graphical-based interface, we can conclude that the images have to be colored drawings (see Fig 11), with a shape close to the real object, but not necessarily similar to the database image (see Fig 10). Moreover, 75% of the users preferred a big icon without text to a smaller icon with the object name (see Fig 9). They affirmed that "A picture is worth a thousand words" and this was especially for the elderly suffering of visual impairment.

Concerning the consistency of the proposed images, we

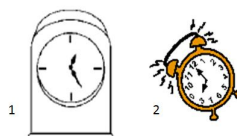


Fig. 10. For the function WakeUp 1. Database image; 2. Preferred image for the robot graphical-based interface.

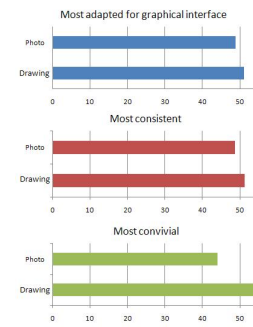


Fig. 11. Comparison between photos and drawings chose as most adequate for the interface, most consistent and convivial.

obtained a very strange result: only 21% of interviewed people have chosen the database of images as the most consistent (strange because the images of database have been considered the best representation of the objects). Much more expected is the fact that only 4.2% of people have chosen these images as the most convivial and only 10% of people would have these images on the robot graphical-based interface. In fact, the database of images is very simple and composed of black-and-white drawings. 51.33% of people have chosen a drawing as the most consistent image (48.67% have chosen a photo). For some robot functions, we have proposed the choice between an image (drawing or photo) and an icon type button. The following results have been obtained:

- Only 16% of users think that the image type button is more consistent than a photo or a drawing
- 32% believe that a button is more convivial
- And 16% prefer an image type button to a photo or a drawing for the robot interface

Regarding the usability of the 2 proposed keyboards layouts (azerty and alphabetical), the performances of all patients were much better with the azerty keyboard layout. This is explained by the fact that all patients who participated to the test have a computer at home and use it regularly. The average time to write 3 words ("Bonjour"¹¹, "Soleil"¹², and "Village"¹³) on the robot touch screen gave an average of 6s (average per word) with the azerty keyboard layout and 13.4s with the alphabetical keyboard layout. Patients were very destabilized by the alphabetical keyboard layout (see Fig 13).

We have also learnt from the medical staff that for the patients who do not have the habit of using computers, it is much better to use alphabetical keyboard layouts. For these patients (especially in the cases where the patients suffer from cognitive impairments) the classic azerty keyboard layout is very difficult and dissuading. For these reasons, we conclude that our system should offer to the user the possibility to choose the keyboard layout according to their

¹¹" Good morning "

¹²" Sun "

¹³" Village "



Fig. 12. Example of robot interface that we are developing following the results of the graphical-based interface.

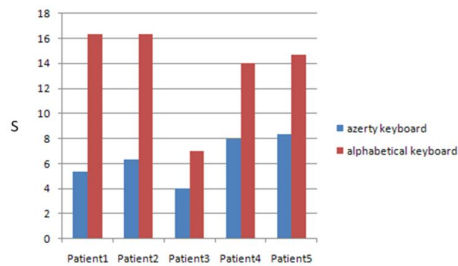


Fig. 13. The average time to write the three French words "Bonjour", "Soleil", and "Village" with the two types of keyboards for each participant.

preferences and habits.

VII. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed the study of two human-robot interfaces - the voice-based and the graphical-based interfaces - to interact with an assistive robot designed for the elderly. Our work mainly focuses on the fission process of these two interfaces. We showed that the choice of the vocabulary, of the syntactic models, and of the images is very delicate and important so as to improve the interaction between the user and the robot. The words and the sentences used by the elderly don't usually match with the language of non-elderly adults people. The images preferred for the graphical-based interface are not the images known in the literature as the most consistent. This means that the robot has to adapt its communication channel to the elderly and the interfaces have to be customized for this type of users. Future work will first focus on the modification of the existing voice-based interface with the syntactic structures and vocabulary based on the results obtained from these tests and described in this paper. We are also developing a graphical-based interface using the images selected by patients (see an example in Fig 12). These improvements will help us to demonstrate that the processing of a target word or of a target picture is facilitated if the sound stimulus (voice) and the visual stimulus (images) are presented at the same time. In other words, if the interface between the robot and the patient is implemented on two different modalities

(i.e., voice and graphical-based interfaces) the accessibility of the machine increases considerably.

REFERENCES

- [1] M. K. Gusmano, Victor G. Rodwin, "The Elderly and Social Isolation". Testimony to Committee on Aging, NYC Council, February 13, 2006.
- [2] C. Victor, S. Scambler, J. Bond and A. Bowling, "Being Alone in Later Life: Loneliness, Social Isolation and Living Alone". *Reviews in Clinical Gerontology*, 2000, pp. 407-417.
- [3] C. Bartneck, J. Forlizzi, "A Design-Centred Framework for Social Human-Robot Interaction". In *Proc. Ro-Man2004*, Kurashiki, 2004, pp. 591-594.
- [4] F.Gödde, S. Mller, K.-P. Engelbrecht, C. Khnel, R. Schleicher, A. Naumann, M. Wolters, "Study of a speech-based smart home system with older users". In *Proc. Int. Workshop on Intelligent User Interfaces for Ambient Assisted Living*, 2008, pp. 17-22.
- [5] T.A. Salthouse, "What and when of cognitive aging". *Current Directions in Psychological Science* 13, 2004, pp. 140-144.
- [6] A. Haasch, S. Hohennner, S. Huwel, M. Kleinhagenbrock, S. Lang, I. Toptsis, G. Fink, J. Fritsch, B. Wrede, and G. Sagerer, "Biron - the bielefeld robot companion". In *Proc. Int. Workshop on Advances in Service Robotics Stuttgart Germany*, 2004, pp. 27-32.
- [7] K. Nickel and R. Stiefelhagen, "Real-time recognition of 3d-pointing gestures for human-machine-interaction". In *Proc. Int. Workshop on Human-Computer Interaction HCI*, Prague, 2004.
- [8] R. Atienza, A. Zelinsky, "Active Gaze Tracking for Human-Robot Interaction". In *Proc IEEE International Conference on Multimodal Interfaces*, 2002, pp. 261-266.
- [9] C. Guo and E. Sharlin, "Exploring the use of tangible user interfaces for human-robot interaction: a comparative study". In *Pro. ACM CHI Conference on Human Factors in Computing Systems*, Florence, 2008.
- [10] K. Severinson-Eklundh, A. Green, H. Httenrauch, "Social and collaborative aspects of interaction with a service robot". *Robotics and Autonomous Systems*, 2003, pp. 223-234.
- [11] P. Rouanet, J. Bchu, P.-Y. Oudeyer, "A comparison of three interfaces using handheld devices to intuitively drive and show objects to a social robot: the impact of underlying metaphors". *Ro-Man 2009*.
- [12] R. Reichman, "Communication Paradigms for a Window System". In: Norman, D.A. / Draper, S.W. (eds): *User Centered System Design: New Perspectives on Human- Computer Interaction*. Hillsdale, NJ & London: Lawrence Erlbaum, 1986, pp. 285-313.
- [13] R. Reichman, "Integrated Interfaces Based on a Theory of Context and Goal Tracking". In: Taylor, M.M. / Neel, F. / Bouwhuis, D.G. (eds): *The Structure of Multimodal Dialogue*. Amsterdam et al.: North-Holland, 1989, pp. 209-228.
- [14] D. Perzawnoski, A. C. Shultz, W. Adams, E. Marsh, M. Bugajska, "Building a Multimodal Human-Robot Interface". *IEEE Intelligent System* 2001.
- [15] M. Wolters, K. Georgila, J. D. Moore, R. H. Logie, S. E. MacPherson, "Reducing working memory load in spoken dialogue systems". *Interacting with Computers*, 2009, pp.276-287.
- [16] J.G. Snodgrass, M. Vanderwart, "A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity". *Journal of Experimental Psychology: Human Learning and Memory*. Vol 6(2) , Mar 1980, pp. 174-215.
- [17] F.-X. Alario, L. Ferrand, "A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition". *Behavior Research Methods, Instruments, & Computers*, 1999, pp. 531-552.
- [18] P. Bonin, R. Peereman, N. Malardier, A. Mot, M. Chalard. "A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies". *Behavior Research Methods, Instruments, & Computers*, 2003, pp. 158-167.
- [19] <http://www.mayer-johnson-symbols.com/aboutsymbols/pcs.htm>
- [20] <http://www.mayer-johnson.com/products/boardmaker/>
- [21] <http://fr.fotolia.com/>
- [22] [http:// leadserv.u-bourgogne.fr/bases/pictures/](http://leadserv.u-bourgogne.fr/bases/pictures/)
- [23] <http://www.lexique.org/telLexique.php>